ARTÍCULOS

MARCADO SEMÁNTICO AUTOMÁTICO EN GESTORES DE CONTENIDOS: INTEGRACIÓN Y CUANTIFICACIÓN

Juan-Antonio Pastor-Sánchez, Enrique Orduña-Malea y Tomás Saorín

Juan-Antonio Pastor-Sánchez es doctor en documentación y profesor de la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia*, en el área de la construcción de servicios y sistemas de información digital. También ha desempeñado su carrera profesional como documentalista y diseñador de sistemas de información, creación de entornos de enseñanza en red y diseño web. Colabora con el *W3C* en la traducción de material de referencia de *SKOS* y otras tecnologías de la web semántica. Investiga la aplicación de las tecnologías de la web semántica, *linked open data*, diseño de ontologías, gestión de contenidos digitales y arquitectura de la información.
http://orcid.org/0000-0002-1677-1059

Universidad de Murcia. Facultad de Comunicación y Documentación
Campus de Espinardo, 30071 Murcia, España
pastor@um.es

Enrique Orduña-Malea es doctor en documentación por la *Universidad Politécnica de Valencia* (*UPV*). Investigador posdoctoral en el *Instituto de Diseño y Fabricación* (*IDF*) y profesor externo en el *Departamento de Comunciación Audiovisual, Documentación e Historia del Arte* (*DCADHA*) en la *UPV*. Desde 2012 es miembro del grupo de investigación *EC3* de la *Universidad de Granada*. Entre otras actividades, es subdirector del *Anuario ThinkEPI* y miembro del equipo del ranking de profesores *H Index Scholar*. Su área de investigación es la cibermetría (tanto descriptiva, instrumental como aplicada), orientada especialmente a la creación, difusión y consumo de información académica en internet y a la cuantificación y visualización de estas actividades.
http://orcid.org/0000-0002-1989-8477

Universidad Politécnica de Valencia, Escuela Técnica Superior de Informática
Edificio 1H. Camino de Vera, s/n. 46022 Valencia, España
enorma@upv.es

Tomás Saorín es profesor asociado en la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia y* documentalista de la *Comunidad Autónoma de Murcia* donde ha participado en la puesta en marcha de proyectos de gestión de contenidos e información institucional en las áreas de servicios sociales, trabajo, empleo y bibliotecas. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha investigado sobre estrategia digital y de edición electrónica, así como sistemas de gestión de contenidos y proyectos colaborativos.
http://orcid.org/0000-0001-9448-0866

Universidad de Murcia. Facultad de Comunicación y Documentación
Campus de Espinardo, 30071 Murcia, España
tsp@um.es

Resumen

Se ofrece en primer lugar una visión general de los diferentes formatos de marcado semántico así como de las tecnologías existentes para incorporar información semántica explícita (microformatos, microdatos y *RDFa*). Posteriormente se describen algunos servicios que permiten automatizar los procesos de anotación semántica (*Sindice*, *Calais*, *AlchemyAPI* y *DBPedia Spotlight*) al tiempo que se caracteriza el ciclo completo de este proceso en un CMS concreto (*Wordpress*) mediante un *plugin* especializado (*RDFaCE-Lite*). Finalmente, con el propósito de cuantificar la creación y la conectividad del contenido marcado semánticamente en la Web, se analiza el conjunto de universidades españolas (y una selección de 25 universidades internacionales) mediante *Sindice*. Para ello se calculan indicadores de tamaño semántico y de enlaces entrantes, salientes, internos y de terceros (*third party links*) en los *datasets* de las universidades de la muestra. Los resultados indican la todavía escasa presencia de contenido marcado semánticamente en las universidades, así como el alto aislamiento en visibilidad web de estos contenidos.

Palabras clave

Web semántica, *Linked data*, Marcado semántico, Gestores de contenidos, *Wordpress*, *RDFaCE-Lite*, *Sindice*, Universidades, Webometría.

Artículo recibido el 15-06-2013
Aceptación definitiva: 18-09-2013

Title: Automatic semantic markup in content management systems: integration and quantification

Abstract

A general overview of the different semantic markup formats and the existing technologies to incorporate explicit semantic information (microformats, microdata and RDFa) is provided. Services are described that automate, to some extent, semantic annotation processes (*Sindice*, *Calais*, *AlchemyAPI* and *DBPedia Spotlight*), while characterizing the complete cycle of this process in a particular CMS (*Wordpress*) using a specialized plugin (*RDFaCE-Lite*). Finally, in order to quantify the creation and connectivity of semantically marked content on the Web, the space formed by all Spanish universities (and a selection of 25 international institutions) is analysed with *Sindice*. Semantic page count and visibility indicators (inlinks, outlinks, internal and third party) are calculated for the sample. The results indicate limited presence of semantically marked content in the universities and highly isolated web visibility of this content.

Keywords

Semantic web, Linked data, Semantic markup, Content management systems, *Wordpress*, *RDFaCE-Lite*, *Sindice*, Universities, Webometrics.

Pastor-Sánchez, Juan-Antonio; **Orduña-Malea, Enrique**; **Saorín, Tomás** (2013). "Marcado semántico automático en gestores de contenidos: integración y cuantificación". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 381-391.

http://dx.doi.org/10.3145/epi.2013.sep.02

1. Introducción

El ecosistema que conforma la web semántica suele tratarse como una propuesta a largo plazo. Actualmente existe un intenso foco de interés en conectar conjuntos de datos (*datasets*), dando lugar a lo que se denomina *linked open data* (datos enlazados). Este fenómeno sigue unas pautas ampliamente conocidas y sugeridas inicialmente por **Tim Berners-Lee**. En la Web se publican grandes conjuntos de datos de una manera estructurada siguiendo dichas pautas, al tiempo que se definen de forma explícita las relaciones entre recursos mediante uno o varios vocabularios, como por ejemplo RDF, OWL (*web ontology language*) o SKOS (*simple knowledge organization system*) entre otros (**Pastor-Sánchez**, 2011). *DBpedia* es el *dataset* más utilizado en el universo *linked open data*.
http://www.w3.org/wiki/LinkedData
http://dbpedia.org

El funcionamiento eficiente de *datasets* interconectados ofrece enormes posibilidades de aplicación práctica, desde la mejora de la recuperación de información en motores de búsqueda generalistas al diseño de servicios de valor añadido en ámbitos de uso intensivo de información, como la documentación científica (**García-García**, 2012) o las bibliotecas (**Sellés-Carot**; **Orduña-Malea**; **Serrano-Cobos**, 2013) entre otros, aunque sus aplicaciones aún están en fase incipiente, pendientes de masa crítica y de *killer apps* (**Saorín**; **Peset**; **Ferrer-Sapena**, 2013).

Es preciso conceptualizar una dualidad, en la que *datasets* y contenidos web deberían dar forma a una realidad en la que los procesos de publicación partieran de la interrelación y la consiguiente integración de grandes cantidades de datos y contenidos informativos.

Los sistemas de gestión de contenidos (CMS) han alcanzado durante la última década un alto grado de desarrollo conceptual desde el punto de vista de la arquitectura de la información, la interconexión con bases de datos corporativas, la modularidad y la reutilización de contenidos. La disponibilidad en abierto de muchas de estas aplicaciones y su facilidad de instalación y manejo han propiciado un uso masivo por parte de los cibernautas.

Según *Builtwith* -a mayo de 2013- los tres CMS más utilizados son *Wordpress*, *Joomla!* y *Drupal*, sumando entre los tres casi 10 millones de instalaciones. Por su parte *W3Tech* apunta que el 58% de los sitios web gestionados mediante CMS utilizan *Wordpress*, mientras que los sitios que no utilizan ningún CMS se reducen un 11% en los últimos 3 años.
http://trends.builtwith.com/cms/top
http://w3techs.com

Esto no hace más que poner de manifiesto el enorme peso que, en la creación y difusión de contenidos web a nivel global, tienen actualmente los CMS, que han ido a su vez evolucionando con el tiempo:

1ª fase: contenido y presentación mezclados;
2ª fase: contenido y presentación separados;
3ª fase: contenido, presentación y significado separados.
http://www.webnodes.com/why-semantic-cms

Aun así, los CMS tienen todavía una escasa participación en el ciclo de vida (creación, publicación y reutilización) de los *datasets* enlazados y, en consecuencia, existe un bajo de nivel de integración de éstos con los contenidos web (**Pastor-Sánchez**, 2012). Dicha integración supone un reto y una necesidad, puesto que actualmente el uso cada vez mayor conlleva una "larga cola" en cuanto a la producción de contenidos web, que aporta una diversidad informativa insustituible.

Dadas las facilidades que los CMS aportan en la gestión de sitios web (diseño, usuarios, contenidos, servicios, etc.), resulta fundamental que éstos se expandan y participen igual-

mente en el ciclo de vida de la gestión de conjuntos de datos a partir del marcado semántico de contenidos, analizado más adelante. La clásica idea de una web para máquinas y una web para personas encuentra un punto de convergencia en este enfoque. Se trata de un flujo en el que contenidos y conjuntos de datos intercambian información.

La extracción de información semántica de los propios contenidos web, publicados de forma extremadamente distribuida, permitiría mejorar la eficiencia de los motores de búsqueda y ofrecería un mecanismo para la creación automática de conjuntos de datos semánticos. Por otro lado, el marcado semántico abre nuevas posibilidades para la reutilización de *datasets* como fuente de información semántica que sería incorporada a los contenidos.

Este trabajo muestra y analiza las posibilidades que aportarían los procesos de marcado semántico en la gestión de contenidos mediante CMS. Asimismo se utilizan ciertos indicadores web para determinar el nivel de desarrollo y aplicación del marcado semántico.

2. ¿Qué es el marcado semántico de contenidos web?

El marcado semántico de contenidos supone la incorporación explícita de información a documentos web, de forma que éstos sean inteligibles y procesables por parte de aplicaciones informáticas. Básicamente se trata de añadir determinados atributos (de un modo transparente al lector) al marcado (x)html, con la finalidad de identificar tanto objetos como propiedades de los mismos, así como las relaciones entre ellos.

2.1. Formatos de marcado semántico

Con independencia del esquema de descripción utilizado, los principales formatos usados en el marcado semántico son microformatos, *RDFa* y microdatos.

Microformatos

La técnica de los microformatos (**Khare**; **Çelik**, 2006) se basa en el uso de xhtml (versiones 1 y 5) y html (versiones 4 y 5) de forma combinada con css. En esencia se utiliza el atributo "class" para identificar vocabularios, clases y propiedades. La clave de su éxito inmediato y rápida difusión se encuentra en el uso de tecnologías ya existentes, y en la adopción de una serie de convenciones muy sencillas para referirse a los vocabularios y elementos de descripción utilizados. Sin embargo, uno de sus principales problemas es la ambigüedad producida por la combinación, en un mismo atributo "class", de la información relativa a los estilos de formato visual y a la descripción del contenido semántico de un elemento. Tampoco es posible utilizar URIs para identificar los elementos sobre los que se realiza la descripción y no existe una correspondencia clara con RDF, que entorpece la extracción de información interoperable.

RDFa

La solución del *W3C* para el marcado semántico es *RDFa*, fundamentada en una total compatibilidad con la tecnología clave de la web semántica: RDF. Su última versión (*RDFa* 1.1) utiliza un conjunto de atributos específicos y la reutilización de otros ya existentes para la inclusión de información semántica.
http://www.w3.org/TR/rdfa-syntax

La especificación contempla su uso en diferentes lenguajes "anfitrión" como xhtml (versiones 1 y 5), html (versiones 4 y 5), xml, SVG (*scalable vector graphics*), ePUB y *OpenDocument*. La potencia de *RDFa* contrasta con cierta complejidad en su aplicación. Su creación ha precisado de la definición de nuevos atributos y la reutilización de otros. Por este motivo, el *W3C* desarrolló la recomendación *RDFa Lite*, que indica cómo utilizar *RDFa* de un modo más simplificado, cubriendo las necesidades más comunes de marcado semántico.
http://www.w3.org/TR/rdfa-primer
http://www.w3.org/TR/rdfa-lite

Podría decirse que *RDFa* es una alternativa flexible al tiempo que potente para la inclusión en una página web de cualquier tipo de sentencia RDF.

Microdatos

La tercera opción para el marcado semántico son los microdatos, cuyo desarrollo está ligado al de html5. El grupo de trabajo *Whatwg* se creó en 2004 cuando el *W3C* decidió apostar por las tecnologías de marcado basadas en xml, como xhtml, abandonando html. El objetivo de dicho grupo era la creación de un lenguaje de marcado con nuevos elementos de marcado estructural, de fácil aplicación en dispositivos con baja disponibilidad de recursos y susceptible de integrar capacidades de reproducción de contenido multimedia.
http://www.whatwg.org

Actualmente los trabajos de estandarización de html5 están coordinados por el *W3C*, puesto que este organismo decidió no continuar con el desarrollo de xhtml2. Se decidió que la especificación de html5 podría utilizarse con dos serializaciones: una basada en una DTD (html5) y otra basada en xml (xhtml5). Los microdatos son la solución nativa de html5 para incorporar contenido semántico. Además, el proyecto *Schema.org*, liderado por *Google*[1], ha supuesto un espaldarazo a este formato de marcado semántico. La propuesta define propiedades nuevas (aunque menos que en el caso de *RDFa*) y reutiliza otras ya existentes.
http://schema.org
http://www.w3.org/TR/microdata

El más potente de los tres formatos es sin duda *RDFa*, pero también es la opción más compleja: su aplicación implica cierto dominio y comprensión de RDF, e incluso de OWL si se desea una estructuración formal de los datos semánticos incorporados en el marcado. En cuanto a la formalización semántica, los microformatos representan la opción con mayores carencias. Los microdatos ofrecen una solución in-

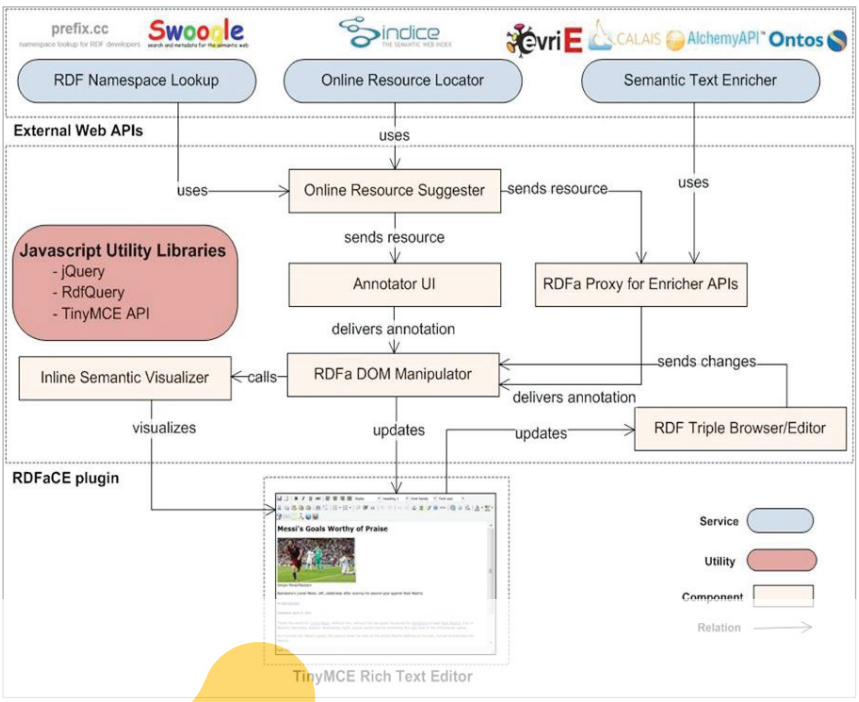

Figura 1. Arquitectura funcional de *RDFaCE* (fuente: **Khalili**; **Auer**; **Hladky**, 2012)

termedia que pueden utilizarse con la mayoría de vocabularios RDF, además de *Schema.org*.

2.2. Marcado semántico y CMS

Los CMS son herramientas de gran ayuda en el proceso de marcado semántico en cualquiera de los tres formatos. Incluso es posible combinar los tres formatos dentro de un mismo documento, aumentando así las posibilidades de procesamiento automático de tales contenidos. Puesto que el *backend* de un CMS genera el código (x)html, el administrador de un sitio web únicamente deberá definir las equivalencias entre las estructuras de contenido con los elementos de uno o varios vocabularios de descripción. A partir de ahí, el propio CMS se encarga de realizar el marcado adecuado. Son numerosas las extensiones en distintos CMS que implementan esta funcionalidad de un modo sencillo y muy escalable. Aquí cabe destacar la potencia de *Drupal* (a partir de su versión 7) que incorpora el marcado semántico mediante *RDFa*, tanto en su núcleo como mediante extensiones. El tipo de información susceptible de este enfoque de marcado semántico es información muy estructurada: datos de personas, productos, libros, películas e incluso recetas de cocina.

http://www.w3.org/wiki/Mixing_HTML_Data_Formats
https://drupal.org/project/rdfx

Pero no hay que olvidar que la mayor parte de los contenidos web no están sujetos a estructuras tan definidas. La mayor parte de los CMS siguen organizando los contenidos usando un conjunto de metadatos descriptivos (título, autor, fecha) asociados a un campo de texto principal sobre el que se pueden aplicar distintos tipos de formatos, definir enlaces o insertar imágenes. Sin embargo, es precisamente en el cuerpo del contenido web donde se encuentra realmente el mensaje en sí mismo. Por tanto, la cuestión es ¿qué sucede con el texto ubicado en el clásico campo "body" desde el punto de vista del marcado semántico?

2.3. Servicios de datos semánticos para el marcado semántico automático

Existen algunas herramientas que amplían las funciones de los CMS para el marcado semántico manual del cuerpo de los contenidos. Pero realizarlo manualmente resulta tedioso, poco operativo e incluso puede ocasionar un marcado deficiente y con errores. Existen proveedores de servicios que facilitan este trabajo, sugiriendo e incluso realizando directamente el marcado. Su uso suele realizarse a través de APIs (*application programming interfaces*) que permiten su portabilidad a múltiples plataformas. Los CMS se conectan con estos servicios mediante extensiones que hacen de puente entre dichas APIs y el motor de marcado de contenido, generalmente basado en un editor *wysiwyg*. Las APIs devuelven los resultados en formatos como xml, RDF o json (*javascript object notation*).

AlchemiAPI

Se trata de un saas[2] basado en técnicas de procesamiento del lenguaje natural (PLN). Se utiliza mediante una API que interactúa con un servicio web o un *kit* de desarrollo disponible en múltiples lenguajes de programación. Esta API es en realidad una interfaz que permite la explotación de bases de datos estadísticas y lingüísticas mediante un conjunto de complejos algoritmos. Tras suministrar un texto, en alguno de los idiomas soportados, es posible identificar personas, organizaciones, empresas, conceptos, categorías, extraer diferentes tipos de entidades incluso pertenecientes a conjuntos de datos LOD externos.

http://www.alchemyapi.com

> Con independencia del esquema de descripción utilizado, los principales formatos usados en el marcado semántico son: microformatos, *RDFa* y microdatos

Sindice

Este servicio en línea básicamente es un motor de búsqueda que recopila documentos RDF y los indiza con el objeto de facilitar la búsqueda de recursos (**Oren** *et al.*, 2008). Localiza documentos que contengan sentencias sobre un determinado recurso. En el marcado semántico resulta útil para definir sentencias sobre dichos recursos a partir de las entidades identificadas en el contenido web. *Sindice* indiza documentos web que utilicen cualquier tipo de formato para el marcado semántico y dispone de APIs para la búsqueda y la consulta de los índices, así como de herramientas para la extracción de tripletas RDF y la validación del marcado.

http://www.sindice.com

OpenCalais

Es un servicio web de *Thomson-Reuters* que también utiliza técnicas de PLN para el enriquecimiento semántico de contenidos web. El funcionamiento sigue la misma pauta que en *AlchemiAPI*: se envía un texto al servicio web y éste, tras analizarlo, localiza entidades, hechos y eventos. Ésta es una diferencia significativa con *AlchemiAPI*, que únicamente se limita a la identificación de entidades. En principio este servicio gratuito está limitado a 50.000 transacciones diarias. Su uso de la API puede realizarse mediante invocaciones a servicios web soap (*simple object access protocol*) o rest (*representational state transfer*). El servidor responde devolviendo código html marcado con microformatos o declaraciones RDF en los formatos json, N3 (notation3) o *simple format*.
http://www.opencalais.com

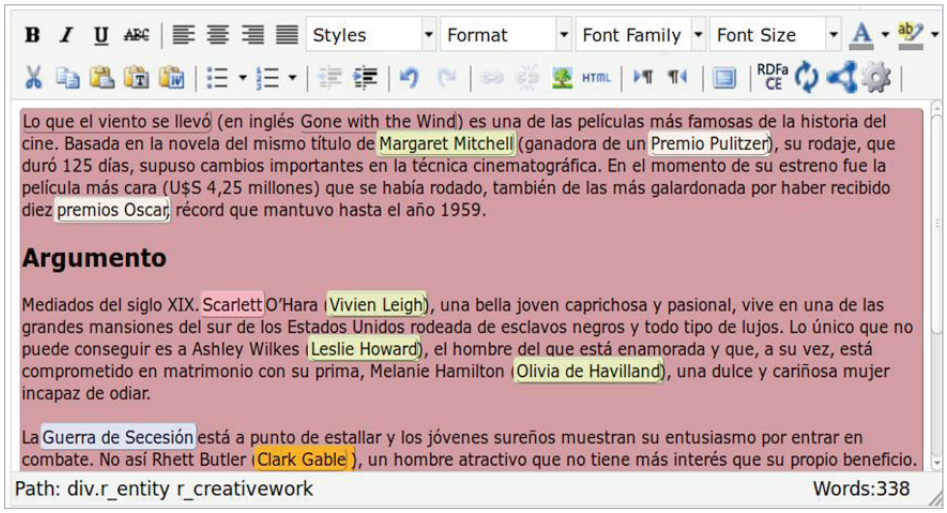

Figura 2. Ejemplo de marcado semántico (manual y automático) de un fragmento de texto usando la instalación demo de *RDFaCE*
http://rdface.aksw.org/new/tinymce/examples/rdface.html
(fuente: *http://rdface.aksw.org*)

DBpedia Spotlight

Este servicio permite la aplicación de la ontología de *DBpedia* en tareas de marcado semántico, basándose en técnicas enfocadas a la desambiguación semántica. Permite identificar entidades de una de las 272 clases de la ontología (**Mendes *et al.***, 2011). *Spotlight* puede usarse a través de un servicio web restful/soap o utilizando el archivo java ofrecido por este proyecto para desarrollar una aplicación web. Mediante *Spotlight* es posible reconocer entidades, anotar un texto o desambiguar un término.
http://spotlight.dbpedia.org

2.4. Caso de uso: marcado semántico en *Wordpress* mediante *RDFaCE*

Como se ha expuesto anteriormente, los CMS comienzan a permitir, generalmente mediante extensiones, el marcado semántico de estructuras y metadatos, aunque el marcado del cuerpo de texto está todavía menos desarrollado.

En *Drupal*, por ejemplo, se dispone del módulo *Semantic markup editor*. Se trata de un proyecto en plena fase de desarrollo inicial para el marcado semántico del contenido dentro del editor *wysiwyg*. Otros módulos para este CMS son *OpenCalais* y *Alchemy*, que hacen uso de las correspondientes APIs para el marcado automático de los contenidos. Otro CMS muy extendido, *MediaWiki*, dispone de la extensión *Semantic MediaWiki* para definir propiedades y atributos, incorporando semántica a los "wikienlaces". Su uso con otros módulos para formularios o *RDFa*, permite trabajar una wiki como si fuera una base de datos, utilizando un lenguaje de interrogación y generando contenidos dinámicos. Un ejemplo de este tipo de funciones es *Referata*, un servicio de alojamiento de wikis semánticas cuyos datos pueden gestionarse y reutilizarse de forma colaborativa.
https://drupal.org/project/semantic_markup_editor
http://drupal.org/project/opencalais
http://drupal.org/project/alchemy
http://semantic-mediawiki.org
http://www.referata.com

Por su parte, *Wordpress* dispone de un *plugin* denominado *RDFaCE* que combina la edición manual y el marcado automático en el propio cuerpo del contenido de forma interactiva. Dado el predominio de *WordPress* en el mercado web, este *plugin* se analiza a continuación con mayor detalle (figura 1).

RDFaCE realiza el marcado semántico utilizando RDFa y microdatos y aplicando los siguientes vocabularios:

- rNews: esquema de metadatos desarrollado por IPTC (*International Press Telecommunications Council*) para la descripción de recursos relacionados con los medios de comunicación.
  http://www.iptc.org/std/rNews/1.0/specification/rNews_1.0-diagram.pdf
- foaf (*friend of a friend*): permite describir personas, sus actividades y las relaciones con otras personas, organizaciones y objetos.
  http://www.foaf-project.org
- *DBpedia*: ontología para la representación de información extraída de *Wikipedia*.
  http://dbpedia.org
- *Schema*: esquema de ámbito general que permite representar una amplia variedad de recursos: personas, trabajos creativos, organizaciones, eventos, etc.
  http://schema.org

El marcado manual se efectúa seleccionando parte o la totalidad del texto del cuerpo del contenido e identificando el tipo de entidad al que se hace referencia: persona, evento, organización, noticia, lugar, objeto audiovisual, etc. En el texto correspondiente a la entidad identificada, se seleccionan fragmentos más específicos para definir sus propiedades: nombre, URL, fechas, etc. También es posible establecer relaciones entre diferentes entidades (figura 2).

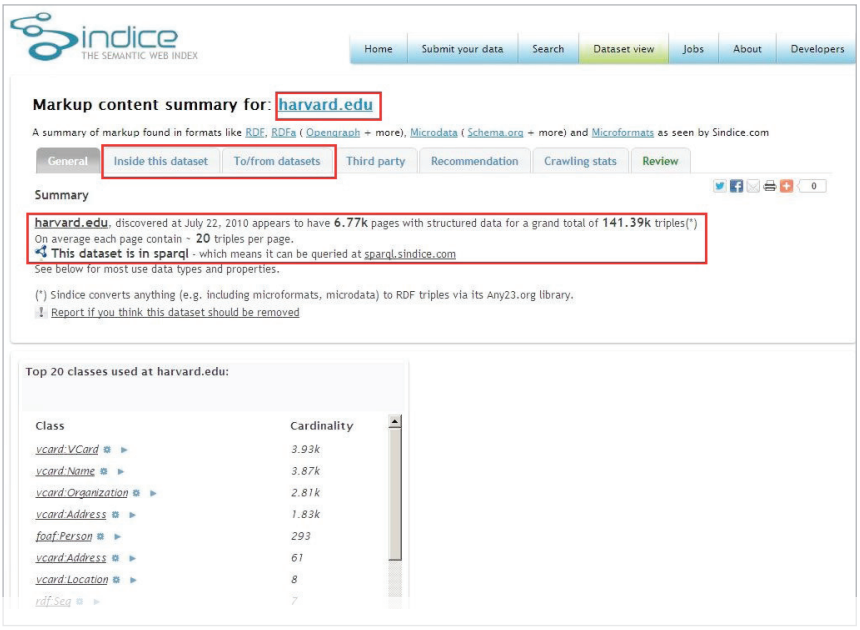

Figura 3. Análisis de *datasets* en dominios web mediante *Sindice* *http://sindice.com*

La auténtica potencialidad de *RDFaCE* puede verse en el uso que hace de tres tipos de servicios semánticos:

- Buscador de espacios de nombres RDF: utiliza *Swoogle* y *Prefix.CC* para identificar los espacios de nombres más adecuados para hacer referencia a las entidades identificadas.
- Localizador de recursos online: busca recursos web a partir de los espacios de nombres identificados previamente, y utiliza los URLs encontrados para marcar como enlaces el texto de las entidades identificadas.
- Enriquecedor semántico: realiza el marcado semántico automático utilizando alguno de los siguientes servicios: *DBpedia Spotlight*, *Ontos*, *OpenCalais*, *AlchemyAPI*, *Extractiv*, *Evri*, *Lupedia* o *Saplo*.

Es necesario puntualizar que *RDFaCE* debe ser configurado para poder utilizar algunos de estos servicios mediante una clave para el uso de la API correspondiente (por ejemplo en el caso de *OpenCalais*).

3. Análisis métrico de la información semántica web

La creación, evolución y extensión del marcado semántico del contenido web, dada su estructuración estandarizada, puede cuantificarse mediante el uso de ciertos indicadores. Esto supone un campo incipiente de análisis de la cibermetría, orientada al análisis métrico de información semántica en la Web.

Hasta la fecha los estudios relativos a la aplicación de técnicas cibermétricas a *linked data* son escasos. **Longqing** y **Qingfeng** (2011) estudian el uso de metadatos *Dublin Core* (DC) como fuente de datos para la cibermetría, analizando especialmente la existencia de los elementos "source" o "relation" (que permiten explicitar el tipo de relación entre recursos enlazados mediante URLs). Sin embargo, los autores concluyen que estos metadatos no son suficientes por sí mismos y reclaman la necesidad de publicar mediante patrones de expresión estándar así como una mayor gestión de la organización de recursos web. La existencia de contenidos no estructurados provoca una situación caótica, en la que los robots de los motores de búsqueda no pueden proporcionar información en las condiciones que precisa la cibermetría: la búsqueda directa en etiquetas genera un enorme ruido documental que anula la validez de cualquier análisis estadístico.

El creciente uso de los diversos formatos de marcado semántico explicados anteriormente permite, de una manera más precisa, estructurar las relaciones entre recursos web. Esto ha permitido desarrollar diversas iniciativas para añadir información semántica a los enlaces cuando se aplican en diversos indicadores. Así como el *PageRank* y otros algoritmos similares como *Hits* (**Kleinberg**, 1999) o *SimRank* (**Jeh**; **Widom**, 2002) no tienen en cuenta los tipos de relaciones que los enlaces están proporcionando, otras iniciativas como *ObjectRank* (**Balmin**; **Hristidis**; **Papakonstantinou**, 2004) o *TripleRank* (**Franz** *et al.*, 2009) sí permiten explicitar las relaciones semánticas proporcionadas por los enlaces.

Por otro lado, se han llevado a cabo estudios pioneros en la elaboración de rankings de universidades basados en la información existente sobre éstas en determinados *datasets*. Por ejemplo, **Meymandpour** y **Davis** (2013) utilizan la información estructurada depositada en *DBpedia* para diseñar un ranking mundial de universidades a partir de los datos contenidos en las relaciones "dbo:", logrando correlaciones significativas ($r = 0,85$; $p < 0,01$) con el *Ranking de Shanghai* (*Academic Ranking of World Universities*). *http://www.arwu.org*

La web semántica constituye una fuente fundamental de información en entornos académicos, desde un punto de vista métrico. **Stuart** (2012) analiza la funcionalidad de 5 fuentes de información semántica (seleccionando finalmente *Sindice*) para medir el uso de *foaf* en el seno del espacio académico británico. Para ello, recopila una serie de nuevas métricas web: número de documentos con al menos una determinada propiedad o relación (en este caso "foaf:name"), número total de declaraciones RDF o presencia de ciertos predicados dentro del *dataset* analizado. Estos indicadores se aplican a toda la cobertura de *Sindice* o a un determinado dominio web (por ejemplo una universidad), en el más puro estilo cibermétrico clásico.

No obstante, la aplicación de análisis métricos en el contexto de la web semántica es todavía escasa, así como el estudio y propuesta de nuevas fuentes e indicadores. Para suplir esta carencia, en esta última parte del artículo se propone la realización de un sencillo estudio con el que se pretende:

a) Analizar las prestaciones de *Sindice* como fuente de información para estudios cibermétricos.

Tabla 1. Tamaño semántico y visibilidad web en *datasets* en universidades internacionales (*Sindice*)

Universidad	Dominio	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Harvard University	harvard.edu	6.770	8.660	0	0	6
Stanford University	stanford.edu	13.350	3.020	4.970	111	569
Massachusetts Institute of Technology	mit.edu	10.840	2.820	350	140	204
University of Michigan	umich.edu	185	149	0	0	0
University of Pennsylvania	upenn.edu	206	184	6	0	1
University of California-Los Angeles	ucla.edu	2.170	1.610	0	0	0
University of California-Berkeley	berkeley.edu	266	2.020	0	0	17
Columbia University New York	columbia.edu	643	5.950	0	0	13
Cornell University	cornell.edu	290	2.220	0	0	1
University of Minnesota	umn.edu	312	179	0	0	0
Pennsylvania State University	psu.edu	3.100	2.460	0	0	0
University of Texas Austin	utexas.edu	690	651	1	0	3
Yale University	yale.edu	892	377	18.760	0	19
University of Cambridge	cam.ac.uk	505	1.980	0	0	29
Caltech	caltech.edu	2.510	963	0	0	0
University of Oxford	ox.ac.uk	8.310	13.030	2	1.730	99
Duke University	duke.edu	1.510	841	0	0	1
University of Wisconsin Madison	wisc.edu	293	1.150	0	0	0
Universidade de São Paulo (USP)	usp.br	1.260	1	0	0	1
University of Illinois-Urbana Champaign	illinois.edu	13	0	0	0	0
University of North Carolina-Chapel Hill	unc.edu	310	545	16	0	2
University of British Columbia	ubc.ca	9.020	13.570	0	0	2
University of Washington	washington.edu	3.710	2.750	0	0	1
Princeton University	princeton.edu	165	52	0	0	3
University of Utah	utah.edu	60	1.040	0	0	0

Las posiciones de las universidades responden a la posición obtenida en el ranking web, únicamente en tanto que fuente de la que se ha extraído la muestra de estudio

b) Describir indicadores de tamaño y enlazabilidad a nivel de *datasets*.

c) Aplicar los indicadores anteriores a los *datasets* de un conjunto de dominios web.

3.1. Método de trabajo

Tomando como referencia el *Ranking web of universities*, se recogió la totalidad de universidades del sistema español, tanto públicas como privadas (75), y las 25 primeras universidades del ranking mundial. Pese a que ciertas universidades utilizan varios dominios web, para cada universidad sólo se tomó el url recogido en el ranking web pues, según la metodología de este servicio, es el que mejores resultados ofrece para cada institución.
http://www.webometrics.info

La elección de universidades (en lugar de otras organizaciones o productos) se debe a que son instituciones analizadas extensamente en estudios cibermétricos y que, dado su tamaño y misiones fundacionales, son susceptibles de generar grandes cantidades de contenidos web estructurados, semiestructurados y no estructurados.

La muestra de universidades españolas comprende la totalidad del sistema universitario español, para conocer de manera general y comparada su rendimiento. En el caso de las universidades internacionales, la muestra recoge únicamente las 25 instituciones con mayor cantidad de documentación en abierto (tamaño de su web) y con una mayor visibilidad (enlaces externos recibidos), de manera que los datos tengan la escala (en volumen) adecuada para ser representativos.

Los datos de las 100 universidades (75 españolas y 25 internacionales) se tomaron durante la última semana de mayo de 2013 directamente de *Sindice*, a partir del servicio *Dataset view*. Este servicio permite detectar, para cada dominio web analizado, la cantidad de páginas con marcado semántico en cualquier formato y que en este trabajo se denomina "tamaño semántico" (*semantic page count*). Se indica igualmente el número de tripletas existentes en el total de páginas (figura 3).
http://demo.sindice.net

Por otro lado, se recogen los datos de los siguientes indicadores de enlaces:

a) Enlaces internos en *datasets*: número de enlaces donde el sujeto y el objeto de la tripleta pertenecen al mismo dominio.

b) Enlaces de terceros (*third party*) en *datasets*: número de

enlaces, creados por el dominio web analizado, pero donde el sujeto y/o el objeto de la tripleta no pertenecen a este dominio.

c) Enlaces entrantes a *datasets*: número de enlaces dirigidos hacia algún recurso de un *dataset* albergado en un determinado dominio web.

d) Enlaces salientes de un *dataset*: número de enlaces que salen desde algún recurso de un *dataset* albergado en un determinado dominio web.

Los enlaces de terceros no se deben confundir con los *external links* usados en cibermetría clásica. Este indicador cuantifica si dentro del *dataset* de un dominio determinado se ha creado una declaración RDF donde el objeto o predicado (o ambos) es un URL que no pertenezca al dominio estudiado.

Los 5 indicadores utilizados (tamaño semántico, enlaces internos, de terceros, entrantes y salientes) no se han aplicado hasta la fecha en ningún estudio cibermétrico y, por tanto, suponen una novedad en este tipo de trabajos.

Los valores obtenidos son de utilidad para conocer directamente el rendimiento (en tamaño y visibilidad web) de los *datasets* analizados y, de manera indirecta, podrán ser utilizados para cuantificar el efecto de la adición de contenido semántico desde CMS específicos, debido a la posibilidad de conocer el origen de los enlaces a recursos alojados en determinados *datasets*.

3.2. Resultados del análisis cuantitativo

Los datos obtenidos no pretenden en ningún momento mostrar un ranking de universidades, sino ofrecer el tamaño y visibilidad (medida a través de hiperenlaces) de las universidades con mayor impacto en la Web. Los datos obtenidos para las 25 universidades internacionales se muestran en la tabla 1, mientras que las 75 universidades españolas se ofrecen en la tabla 2.

En cuanto a las universidades internacionales, los resultados relativos al tamaño semántico indican datos dispares y en general discretos. Sólo 5 universidades superan las 10.000 páginas con marcado semántico (*Harvard*, *Stanford*, *MIT*, *Oxford*, *British Columbia*). El resto muestra resultados prácticamente vacíos si se comparan con los tamaños documentales totales de estas universidades (*Harvard*, por ejemplo, supera los 16 millones de documentos indizados por *Google*).

Los indicadores de enlaces muestran asimismo resultados muy bajos. Los datos indican tasas de enlazado interno bajas (a excepción de algunas universidades concretas, como *Oxford*, *British Columbia* o *Harvard*) y tasas de enlazado de terceros prácticamente inexistentes (con la excepción de *Stanford* y, especialmente *Yale*). Respecto a los enlaces entrantes y salientes, la práctica totalidad de los resultados son muy bajos o directamente inexistentes, destacando únicamente *Oxford* en número de enlaces entrantes.

En las universidades españolas se observan patrones similares: cierta aleatoriedad en los datos, una lógica mayor presencia de enlaces internos respecto a los de terceros y, especialmente, los bajos niveles generales de enlaces entrantes y salientes. Hasta en 10 universidades (todas ellas privadas) no se obtienen datos (es decir, no están siquiera recogidos en el catálogo de URLs de *Sindice*). Y en las 65 restantes, 24 no contienen ninguna página con contenido marcado semántico. Entre las 41 universidades con al menos 1 página con contenido marcado semántico, destacan especialmente los niveles logrados para la *Universidad Politécnica de Madrid* (58.530), reflejo directo de la actividad de determinados grupos de investigación, la *Universitat de Girona* (9.560) y la *Universidad de Oviedo* (4.630).
http://www.oeg-upm.net

En el caso de los enlaces internos, destacan la *Universitat de Girona* (52.790) e, inesperadamente, la *Universitat Abat Oliva* (2.050), aunque los resultados son en general bajos o inexistentes (hasta 34 universidades no tienen ningún enlace interno, y 61 ningún enlace de terceros).

En cuanto a los enlaces salientes a *datasets*, los datos son prácticamente inexistentes: sólo 15 generan algún enlace saliente, destacando la *Politécnica de Valencia* (315) y la *Universidad de Valladolid* (152), mientras que ninguna universidad española recibe enlaces entrantes a *datasets*. Este resultado muestra el completo aislamiento en visibilidad del ya escaso contenido generado con marcado semántico.

3.3. Conclusiones del análisis cuantitativo

La asimetría entre el número de enlaces entrantes e internos es un indicio de cierta endogamia en la generación de enlaces, al primar relaciones entre recursos alojados en *datasets* de un mismo dominio. Esto se explica porque la mayoría de los enlaces generados sirven para explicitar relaciones verticales entre recursos de una misma colección.

Por otro lado, la ausencia de enlaces entrantes refleja una baja visibilidad web de dichos recursos. Es decir, los recursos web incluidos en los *datasets* no se enlazan entre ellos (ni se reciben enlaces ni se proporcionan), lo que resulta inesperado dado que las pautas de *linked data* promocionan precisamente el enlazado (en este caso con información semántica incluida) entre recursos web. Los resultados indican en cambio cierto aislamiento de los recursos web marcados semánticamente.

Estas bajas tasas de presencia (74,6% de las universidades españolas no alcanzan los 100 documentos marcados semánticamente) y visibilidad (ninguna universidad española recibe un solo enlace externo) de los recursos web pueden ser un reflejo de la situación descrita en la introducción: los *datasets* están siendo creados en determinados lugares y bajo determinados proyectos específicos. Sin embargo, el hecho de que no se estén generando contenidos marcados semánticamente desde CMS (las aplicaciones generalmente utilizadas para crear sitios web en la actualidad) puede provocar que la cantidad y enlazabilidad de los actuales recursos web alojados en *datasets* sean muy limitados. Este aspecto es de especial importancia en dominios web tan complejos como el de las universidades, donde conviven multitud de CMS diferentes, aunque esto resulte opaco para los usuarios.

No obstante, la muestra analizada es muy reducida y los resultados deberán complementarse en el futuro con muestras más amplias de instituciones, así como de otras organizaciones.

Tabla 2. Tamaño semántico y visibilidad web en *datasets* en las universidades españolas (*Sindice*)

Universidad	Url	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Universidad Complutense de Madrid	ucm.es	1	374	0	0	6
Universitat Politècnica de Catalunya	upc.edu	600	1.150	0	0	1
Universidad Politécnica de Madrid	upm.es	58.530	22	0	0	0
Universitat Autònoma de Barcelona	uab.cat	137	57	0	0	0
Universidad de Granada	ugr.es	816	937	0	0	0
Universitat de Barcelona	ub.edu	3.780	2.840	0	0	8
Universidad de Valencia	uv.es	61	497	0	0	0
Universidad Politécnica de Valencia	upv.es	331	4.160	19	0	315
Universidad de Zaragoza	unizar.es	0	0	0	0	39
Universidad de Sevilla	us.es	871	660	3	0	18
Universidad Autónoma de Madrid	uam.es	3	28	0	0	5
Universidad del País Vasco	ehu.es	48	9	0	0	0
Universidad de Alicante	ua.es	111	247	0	0	0
Universidad de Vigo	uvigo.es	1	0	0	0	0
Universidad de Salamanca	usal.es	0	0	0	0	0
Universidad de Navarra	unav.es	18	0	0	0	0
Universidad de Santiago de Compostela	usc.es	0	0	0	0	0
Universitat Pompeu Fabra	upf.edu	2	5	0	0	1
Universidad de Murcia	um.es	139	433	0	0	4
Universidad de Málaga	uma.es	0	0	0	0	0
Universidad de Valladolid	uva.es	181	822	0	0	152
Universidad de Castilla la Mancha	uclm.es	1	0	0	0	0
Universitat Jaume I	uji.es	229	222	0	0	1
Universidad Carlos III de Madrid	uc3m.es	0	0	0	0	0
Universidad de Córdoba	uco.es	315	6	0	0	0
Universitat de Girona	udg.edu	9.560	52.790	3	0	0
Universidad de La Coruña	udc.es	78	108	0	0	0
Universidad de Cádiz	uca.es	324	1	0	0	0
Universitat de les Illes Balears	uib.es	2	4	0	0	0
Univ. Nacional de Educación a Distancia	uned.es	1	0	0	0	0
Universidad de Oviedo	uniovi.es	4.630	188	0	0	2
Universidad de León	unileon.es	0	0	0	0	0
Universidad de Las Palmas de Gran Canaria	ulpgc.es	1	3	0	0	0
Universitat Oberta de Catalunya	uoc.edu	52	72	2	0	2
Universidad de Huelva	uhu.es	0	0	0	0	0
Universidad de Alcalá	uah.es	1	4	0	0	1
Universidad de Cantabria	unican.es	0	0	0	0	0
Universidad de Jaén	ujaen.es	0	0	0	0	0
Universidad de Extremadura	unex.es	0	0	0	0	0
Universidad de La Laguna	ull.es	2	0	0	0	0
Universidad Rey Juan Carlos	urjc.es	0	0	0	0	0
Universitat Rovira i Virgili	urv.cat	55	49	0	0	0
Universidad Miguel Hernández	umh.es	0	0	0	0	0
Universidad Pablo Olavide	upo.es	0	0	0	0	0
Universidad de Almería	ual.es	0	0	0	0	0
Universidad Politécnica de Cartagena	upct.es	1	0	0	0	0
Universidad Pública de Navarra	unavarra.es	0	0	0	0	0
Universidad de La Rioja	unirioja.es	17	0	0	0	0
Universitat de Lleida	udl.es	0	0	0	0	0

Universidad	Url	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Universidad de Deusto	deusto.es	594	1.390	0	0	1
Universidad Pontificia de Comillas	upcomillas.es	0	0	0	0	0
Universitat Ramon Llull	url.edu	0	0	0	0	0
Universidad de Burgos	ubu.es	0	0	0	0	0
Universitat de Vic	uvic.cat	4	4	0	0	0
Universidad Europea de Madrid	uem.es	Sin datos				
Universidad de Mondragón	mondragon.edu	152	0	0	0	0
Universidad Católica San Antonio de Murcia	ucam.edu	220	183	0	0	0
Universidad CEU Cardenal Herrera	uchceu.es	Sin datos				
Universidad San Pablo CEU	uspceu.es	0	0	0	0	0
Universidad Pontificia de Salamanca	upsa.es	2	0	0	0	0
Universitat Internacional de Catalunya	uic.es	0	0	0	0	0
Universidad Internacional de Andalucía	unia.es	299	0	0	0	0
Universidad Nebrija	nebrija.com	Sin datos				
Universidad Internacional de La Rioja	unir.net	Sin datos				
Universidad Camilo José Cela	ucjc.edu	0	0	0	0	0
Universidad Alfonso X El Sabio	uax.es	0	0	0	0	0
Universidad Católica de Valencia	ucv.es	Sin datos				
Universidad Francisco de Vitoria	ufv.es	Sin datos				
Universidad Internacional Menéndez Pelayo	uimp.es	4	4	0	0	0
Universidad Europea Miguel de Cervantes	uemc.es	Sin datos				
Universitat Abat Oliva	uao.es	2.100	2.050	0	0	0
Universidad San Jorge	usj.es	0	0	0	0	0
Universidad Católica de Ávila	ucavila.es	Sin datos				
Universidad a Distancia de Madrid	udima.es	Sin datos				
Valencian International University	viu.es	Sin datos				

Las posiciones de las universidades responden a la posición obtenida en el ranking web, únicamente en tanto que fuente de la que se ha extraído la muestra de estudio

4. Conclusiones finales

Es esencial la existencia de tecnologías y servicios asequibles que asistan a los autores web durante el proceso de marcado. Esto permitiría automatizar ciertos aspectos de este proceso y su integración a gran escala en sitios web gestionados por CMS, para su posterior integración en *datasets*, así como su reutilización en otras plataformas y servicios de forma sencilla. En consecuencia, se mejoraría sensiblemente su difusión y recuperación en la Red, sin la necesidad de diseñar y establecer procesos costosos de marcado semántico manual, o inversión en tecnologías complementarias.

La efectividad del marcado automático aún tiene que ponerse a prueba con rigor. Funciona mejor con textos en inglés que en otros idiomas y hay que hacer varios ensayos de marcado, o seleccionar varios servicios, para obtener un marcado óptimo. Nuestra experiencia sugiere que en última instancia siempre debe recurrirse a una combinación del marcado manual y automático.

Aunque el marcado semántico de textos en un blog no ofrece una ventaja evidente o directa al editor de un web, su incorporación a los CMS más extendidos ayudará a hacer crecer un "excedente semántico" en la Web que sea aprovechable a partir de cierto volumen por otros agentes para producir valor. "Los excedentes grandes son diferentes de los pequeños" (**Shirky**, 2010), y para alcanzar esa masa crítica es necesario reducir el coste de entrada a la larga cola de la web semántica a través de las herramientas más usadas. Dicho contenido semántico también es susceptible de integrarse con *datasets* externos más amplios y, por tanto, en el ecosistema *linked data*.

Estas prácticas facilitarían el desarrollo de estudios cibermétricos que usarían la información semántica introducida en el marcado para obtener nuevos indicadores (como los utilizados en este trabajo para conocer el tamaño y visibilidad web de universidades), y que se pueden expandir a niveles más específicos (especialmente atributos y clases). En ese sentido, *Sindice* ha demostrado ser una fuente útil para la cibermetría, aunque se precisan más estudios para comprobar su cobertura, posibles limitaciones y funcionalidades aplicadas al estudio del crecimiento, evolución y uso de contenido marcado en la Web.

5. Notas

1. También participan en este proyecto *Microsoft* (con su buscador *Bing*), *Yahoo!* y *Yandex*.

2. Saas (*software as a service*): modelo de software en el que las aplicaciones y los datos se ubican en la "nube" y

cuyo acceso y uso se realizan generalmente a través de un navegador web.

6. Bibliografía

Balmin, Andrey; **Hristidis, Vagelis**; **Papakonstantinou, Yannis** (2004). "Objectrank: authority-based keyword search in databases". En: *Procs of the 30th intl conf on very large data bases*, v. 30, pp. 564-575.
http://www.vldb.org/conf/2004/RS15P2.PDF

Franz, Thomas; **Schultz, Antje**; **Sizov, Sergej**; **Staab, Steffen** (2009). "TripleRank: ranking semantic web data by tensor decomposition". En: Bernstein, Abraham *et al.* (ed.). *The semantic web. ISWC 2009*. Springer, v. 5823, pp. 213-228.
http://data.semanticweb.org/pdfs/iswc/2009/paper279.pdf
http://dx.doi.org/10.1007/978-3-642-04930-9_14

García-García, Alicia (2012). *Datos abiertos enlazados linked open data (LOD) en documentación científica*. Valencia: Universidad Politécnica de Valencia.
http://riunet.upv.es/handle/10251/18272

Jeh, Glen; **Widom, Jennifer** (2002) "SimRank: a measure of structural-context similarity". En: *Procs of the 8th ACM Sigkdd intl conf on knowledge discovery and data mining*. New York, pp. 538-543.
http://ilpubs.stanford.edu:8090/508/1/2001-41.pdf
http://dx.doi.org/10.1145/775107.775126

Khalili, Ali; **Auer, Sören**; **Hladky, Daniel** (2012). "The RDFa content editor: from wysiwyg to wysiwym". En: *Computer software and applications conf (Compsac), 2012 IEEE 36th Annual*, pp. 531-540.
http://svn.aksw.org/papers/2012/COMPSAC2012_RDFaCE/public.pdf
http://dx.doi.org/10.1109/COMPSAC.2012.72

Khare, Rohit; **Çelik, Tantek** (2006). "Microformats: a pragmatic path to the semantic web". En: *Procs of the 15th intl conf on world wide web*. ACM: New York, pp. 865-866.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.307.7135&rep=rep1&type=pdf
http://dx.doi.org/10.1145/1135777.1135917

Kleinberg, John M. (1999). "Authoritative sources in a hyperlinked environment". *Journal of the ACM*, v. 46, n. 5, pp. 604-632.
http://www.cs.cornell.edu/home/kleinber/auth.pdf
http://dx.doi.org/10.1145/324133.324140

Longqing, Shi; **Qingfeng, Zhao** (2011). "Data sources of webometrics". En: *7th Intl conf on computational intelligence and security*, pp. 1312-1315.
http://dx.doi.org/10.1109/CIS.2011.291

Mendes, Pablo N.; **Jakob, Max**; **García-Silva, Andrés**; **Bizer, Christian** (2011). "DBpedia spotlight: shedding light on the web of documents". En: *Procs of the 7th Intl conf on semantic systems*, 1-8.
http://goo.gl/JQ2DEs
http://dx.doi.org/10.1145/2063518.2063519

Meymandpour, Rouzbeh; **Davis, Josep G.** (2013). "Ranking universities using *linked open data*". En: *LDOW2013.*
http://events.linkeddata.org/ldow2013/papers/ldow2013-paper-09.pdf

Oren, Eyal; **Delbru, Renaud**; **Catasta, Michele**; **Cyganiak, Richard**; **Tummarello, Giovanni** (2008). "Sindice.com: a document-oriented lookup index for open linked data". En: *Intl journal of metadata, semantics and ontologies*, v. 3, n. 1, pp. 37-52.
http://dx.doi.org/10.1504/IJMSO.2008.021204

Pastor-Sánchez, Juan-Antonio (2011). *Tecnologías de la web semántica*. Barcelona: Editorial UOC, colección El profesional de la información, 1. ISBN: 978 84 9788 474 7

Pastor-Sánchez, Juan-Antonio. (2012). "Los CMS como pieza fundamental en el despliegue de la web semántica". *Anuario ThinkEPI*, v. 6, pp. 184-189.

Saorín, Tomás; **Peset, Fernanda**; **Ferrer-Sapena, Antonia** (2013). "Factores para la adopción de *linked data* e implantación de la web semántica en bibliotecas, archivos y museos". *Information research*, v. 18, n. 1.
http://InformationR.net/ir/18-1/paper570.html

Sellés-Carot, Alicia; **Orduña-Malea, Enrique**; **Serrano-Cobos, Jorge** (2013). "Estrategias y oportunidades tecnológicas en la generación de *linked data* en las bibliotecas". *Mi biblioteca*, pp. 54-59.

Shirky, Clay (2010). *Cognitive surplus: creativity and generosity in a connected age*. New York: Penguin Press. (edición en castellano: *El excedente cognitivo: creatividad y generosidad en la era conectada*. Barcelona: Ediciones Deusto, 2012. ISBN: 978 8423428632). ISBN: 978 0143119586

Stuart, David (2012). "FOAF within UK academic web space: a webometric analysis of the semantic web". En: Widén, Gunilla; Holmberg, Kim (ed.). "Social information research". *Emerald Group Publishing Ltd.*, v. 5, pp. 173-191. ISBN: 978 1 78052 832 8